Collective-Variable Monte Carlo Simulation of DNA

H. A. GABB,[1]* C. PRÉVOST,[1] G. BERTUCAT,[1] C. H. ROBERT,[2] R. LAVERY[1]

[1] *Laboratoire de Biochimie Théorique, CNRS UPR 8090, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, 75005 Paris, France*

[2] *Laboratoire d'Enzymologie et Biologie Structurales, Bâtiment 34-CNRS, 91198 Gif-sur-Yvette, France*

Received 20 September 1996; accepted 20 July 1997

ABSTRACT: Monte Carlo simulations have been carried out on DNA oligomers using an internal coordinate model associated with a pseudorotational representation of sugar repuckering. It is shown that when this model is combined with the scaled collective variable approach of Noguti and Go, much more efficient simulations are obtained than with simple single variable steps. Application of this method to a DNA oligomer containing a recognition site for the TATA-box binding protein leads to striking similarities with results recently obtained from a 1-ns molecular dynamics simulation using explicit solvent and counterions. In particular, large amplitude bending fluctuations are observed directed toward the major groove. Conformational analysis of the Monte Carlo simulation shows clear base sequence effects on conformational fluctuations and also that the DNA energy hypersurface, like that of proteins, is complex with many local, conformational substates. © 1997 John Wiley & Sons, Inc. *J Comput Chem* **18**: 2001–2011, 1997

Keywords: molecular simulation; nucleic acids; internal coordinates; substates; curvature; TATA-box

Introduction

Deoxyribonucleic acids are flexible polymers that fluctuate considerably from their equilibrium structures in solution. These fluctuations allow transient helical bending and twisting, base-pair openings, and conformational adaptation during interactions with other molecules such as proteins and drugs. Consequently, it is important to understand the nature of DNA dynamics to understand its biological activity. Although spectroscopic techniques and, notably, NMR spectroscopy, can yield useful information concerning flexibility it is also necessary to be able to simulate DNA dynamics to interpret these results in terms

* *Present address:* Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK

Correspondence to: R. Lavery

CCC 0192-8651 / 97 / 162001-11

of specific types of conformational fluctuation. (The same requirement applies to interpretation of diffuse X-ray diffraction). Such molecular simulations have generally been performed using molecular dynamics (MD) on oligonucleotide fragments, ideally including a shell of discrete water and counterions to model environmental effects (see, e.g., refs 1–3 for recent examples). Such calculations are, however, computationally demanding and are presently limited to durations of, at most, a few nanoseconds (ns). Shorter *in vacuo* simulations on DNA, using continuum solvent models, have generally led to destabilization of the B-form of the double helix, unless additional *ad hoc* constraints were introduced.

In the present article, we test an internal coordinate model associated with Monte Carlo (MC) sampling as a means for studying DNA fluctuations. Employing internal coordinates, and freezing stiffer degrees of freedom, enables us to reduce the dimensions of the conformational space to be studied by roughly a factor of ten.[4–6] Such an approximation is justified by the study of Amadei et al.,[7] who showed that, in the case of MD simulations on proteins, major movements could be described using only a fraction of the total number of degrees of freedom. This approach can be used in MC simulations if appropriate means are found for dealing with closure problems, such as the treatment of sugar ring puckering. It is, however, less well adapted to MD simulations because of the complex equations of motion that need to be solved.

Although it is also generally assumed that Monte Carlo simulations are less efficient for exploring the conformational space of macromolecules than molecular dynamics,[8] this conclusion is based on a comparison of MC and MD simulations in Cartesian coordinates. The passage to internal coordinates, and the consequent reduction in dimensionality, modifies the situation. Large-scale conformational changes have already been studied in reasonable amounts of computer time using a Monte Carlo algorithm applied to an internal coordinate model requiring a chain closure step.[9,10] We have successfully used a similar system (mixing internal and helicoidal coordinates), Jumna,[11,12] to model controlled deformations in DNA via adiabatic mapping.[13,14] To adapt this approach to Monte Carlo simulations, we have changed to an entirely internal coordinate representation, which avoids the problem of satisfying the backbone closure constraints associated with the use of a mixed internal/helicoidal representation. Satisfying such constraints by frequent minimizations violates the microreversibility condition necessary for statistically meaningful Metropolis Monte Carlo simulations.[15] Moreover, the arbitrary choice of predefined closure points in the backbone and of dependent (and, thus, indirectly sampled) backbone variables is likely to introduce conformational biases.

To improve the efficiency of our Monte Carlo simulations we use a simple transformation of the internal coordinate system that enables DNA to deform along energetically soft pathways in its conformational space, while preserving microreversibility. This is achieved using the scaled collective variable technique developed by Noguti and Go and previously applied to globular proteins.[16] The SCV method uses the Hessian matrix of an initial energy minimized structure to define eigenvectors representing collective movements within the macromolecule. MC steps are taken in these variables with step sizes adjusted, following the corresponding eigenvalues, to insure isotropic movements. We have adapted this method to nucleic acids by employing our recently developed pseudorotational model of sugar flexibility (ref. 17, see also ref. 18), which again avoids a ring closure problem.

In the present article, we describe the basis of this method, demonstrate its superiority to simple MC simulations, and carry out a preliminary analysis of the conformation changes occurring during the simulation of a DNA. The principal tests are made on a DNA 11-mer with the sequence 5′-GTATATAAAAC-3′. This oligomer contains an 8-base recognition site (underlined) for the TATA-box binding protein, TBP.[19] DNA oligomers containing the same sequence have already been studied in our laboratory using both energy mapping with Jumna[20] and molecular dynamics with Amber, during a 1-ns simulation using explicit water and counterions and employing Ewald particle mesh summation to avoid electrostatic truncation.[3] The latter study has shown that this sequence adopts a stable conformation, intermediate between A- and B-DNA, with dominantly C3′-endo sugar puckers for the nucleotides belonging to the TBP binding site. We will term this conformation BAB in what follows. Large bending fluctuations toward the major groove were also observed during the MD simulation. Both these characteristics lead this DNA oligomer to resemble the distorted DNA structure within the cocrystals with TBP.[19]

The clear sequence-dependent behavior of this oligomer made it a good choice for testing our new Monte Carlo approach, which should, in principle, be able to sample large-scale conformational fluctuations. We will show that this is indeed the case. We also analyze our results in terms of the existence of conformational substates, which have been discussed both on the basis of our earlier adiabatic mapping studies,[13,14] and from long MD simulations,[21] with findings similar to those now generally accepted in the case of globular proteins.[22,23]

Last, it should be remarked that although the method we have developed is presently used with a very simple representation of environmental effects, it is compatible with any rapid continuum approach and, indeed, will shortly be combined with a fast Poisson–Boltzmann method (FIESTA), currently under development.[24]

Methodology

NUCLEIC ACID MODEL

DNA conformational freedom is described by the backbone torsions α, β, γ, ε, ζ, and χ; the sugar ring pseudorotational variables phase (P) and amplitude (A); and, in the case of m-stranded systems, by $(m-1)\times 6$ degrees of translational and rotational freedom describing interstrand movement. (The reference axis systems for these variables are placed on the base belonging to the central nucleotide of each strand.) All bond lengths are fixed, as are valence angles, excluding those within the sugar rings. The pseudorotational variables are converted into ring torsions and endocyclic valence angle changes using a recently developed methodology.[17] This approach, which avoids the problem of treating a ring closure condition, involved calculating an optimal energy surface for the sugar ring as a function of its pseudorotational parameters and then developing analytic functions to describe the variations of the intra-cyclic torsion and valence angles over the surface. (Exo-cyclic bonds to the sugar rings are moved in a dependent way with the endo-cyclic valence angle variations to maintain a symmetric position for these substituent atoms). This model may be contrasted with the recent work of Tomimoto and Go who developed an analytic description of ring puckering limited to a single pseudorotational variable.[18]

Using our internal coordinate model, a double-stranded DNA with 12 bp can be represented by only 188 variables (eventually with additional torsions to describe thymine C5-methyl rotation or ribose C2′-hydroxyl rotation in the case of RNA). This number should be compared with more than 2000 variables for standard Cartesian coordinate models.

We calculate conformational energy using the FLEX force field[11,12] described by the following equation:

$$\begin{aligned} E_{Flex} = & \sum_{nb}\left(\frac{q_i q_j}{\varepsilon r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) \\ & + \sum_{\phi}\frac{V_\phi}{2}[1+\cos(n\phi-\gamma)] \\ & + \sum_{\theta} K_\theta(\theta-\theta_{eq})^2 \\ & + \sum_{HB}\cos\theta_{HB}\left[\left(\frac{A_{ij}^{H}-A_{ij}}{r_{ij}^{12}}\right)\right. \\ & \left.-\left(\frac{B_{ij}^{HB}-B_{ij}}{r_{ij}^{6}}\right)\right] \end{aligned}$$

The force field contains nonbonded electrostatic and Lennard–Jones terms, as well as terms describing bond torsions and valence angle deformation. The last term adds an angular dependence to hydrogen bonding interactions, governed by the angle θ_{HB} formed between the vectors A—H and H $\cdots$ B of an A—H $\cdots$ B bond. No nonbonded cutoff distances are used except for the angular dependent hydrogen bond term (9 Å). A full description of this force field and its parameterization may be found elsewhere.[12] Solvent effects are treated using a sigmoidal, distance-dependent dielectric function.[12,25,26]

$$\varepsilon(r) = \frac{(D-D_0)}{D}[(rs)^2+2rs+2]e^{(-rs)}$$

in which D (78) and D_0 (1) are, respectively, the maximum and minimum value of the dielectric at long and short distances and s (0.16) is the slope. Some calculations are also carried out with $D_0 = 4$ to weaken the hydrogen bonding of the DNA base pairs (see Results section). Counterion damping is taken into account by reducing the net charge of the phosphate groups to -0.5e (by adding 0.25e to the punctual charge of each anionic oxygen atom). This simple implicit approach to environmental effects has been shown to be capable of correctly modeling many aspects of nucleic acid conformation and flexibility.[11,26,27]

MONTE CARLO PROTOCOL

The purpose of MC simulations is to generate a Boltzmann ensemble of conformations describing the flexibility of the molecule under study and, eventually, allowing thermodynamic state functions to be calculated. For meaningful results, the conformational space accessible to the molecule at a given temperature must be sampled as thoroughly as possible. For large biomolecules, this represents one of the principal problems facing computer simulations.

Following the standard Metropolis Monte Carlo method,[15] independent variables, q, of the system, chosen at random, are altered by a stepsize Δq. The potential energy is calculated and the new state of the system is accepted if the potential energy difference between the two states, $\Delta E = (E_{i+1} - E_i)$, is negative. Otherwise, the new state is accepted with a probability $e^{(-\Delta E/KT)}$ (where K is the Boltzmann constant and T is the absolute temperature). Modifications of any given variable are made incrementally throughout the simulation, via a probability distribution, $P(\Delta q)$, which is an even function of the stepsize. This distribution is held constant throughout the entire data collection period, so that the simulation satisfies detailed balance conditions.[28] The probability of a given step being accepted depends on the step size. Small steps are generally accepted, but lead to little conformational change. Large steps scan more conformational space, but are likely to be rejected due to associated energy loss. Efficient Monte Carlo simulations aim at acceptance rates ranging between 40% and 50%. We carefully monitor the acceptance level for each variable, while using probability distributions, $P(\Delta q)$, that allow occasional very large moves. The protocol has been applied using two different techniques for changing the variables describing a DNA oligomer.

Single Variable Moves (SV)

In this case, moves are made in the internal variables describing the flexibility of the molecule (backbone torsion and glycosidic angles, sugar phase and amplitude, and interstrand translation and rotation). This choice allows a gain in computer time (with a factor of roughly 2–6) by avoiding the calculation of the internal energy of the parts of the molecule unaffected by the chosen single variable change. We adjust the average stepsize for each independent variable during an equilibration phase in such a way that approximately 50% of the moves of this particular variable are accepted. In this phase, no data are collected other than stepsize and acceptance rate information. When the optimal average step size is determined, the simulation proceeds to the production phase, where energetic and structural data are collected. A test run of 250,000 steps at 300 K was performed on our GTATATAAAAC oligomer in the B conformation, after 200,000 steps of equilibration. We used stepsizes calculated from a uniformly distributed random deviate x in the range $(0 \cdots 1)$:

$$\Delta q_i = \pm 3/2 s_i \ln(x)$$

where s_i is the optimal step size for variable i determined during the equilibration phase. The sign of this Δq_i is chosen via a second random deviate. This equation provides exponentially distributed stepsizes with an average stepsize at $3/2 s_i$, ensuring the possibility of large moves.[29]

Scaled Collective Variables (SCV)

To improve sampling efficiency it is necessary to make large moves. In the case of single variables, this leads to steric clashes and rejection on the basis of energy. This problem could be avoided if it were possible to move along low-energy paths on the conformational energy hypersurface of the macromolecule. A procedure for achieving this has been proposed by Noguti and Go[16] and applied to the study of globular proteins. It consists of moving along directions defined by the eigenvectors of the Hessian matrix describing the molecule in question. We have adapted this method, known as SCV (scaled collective variables), to our studies of DNA.

The shape of the energy surface in the region of a minimum can be approximated by a multidimensional parabola in the space of the conformational variables. The collective variables, ξ_i, describing this surface are defined along the eigenvectors of the Hessian matrix of the surface:

$$\xi = \sum_j u_{ij} \Delta\theta_j$$

where u_{ij} is the jth component of the ith eigenvector ($\sum_j u_{ij}^2 = 1$) and $\Delta\theta_j$ is the deviation of the jth independent variable (backbone dihedral angle, sugar pucker parameter, or interstrand variable) from its local minimum value. For a well-minimized structure the Hessian matrix is positive definite and symmetric. Each eigenvector defines

an independent harmonic variable whose associated force constant is given by the corresponding eigenvalue λ_i. This information can be used to create a new set of scaled collective variables[16]:

$$\eta_i = \lambda_i^{1/2}\xi_i$$

The energy surface in the space of the scaled collective variables is isotropic in the harmonic approximation. This means that the optimal average stepsize will be the same for all collective variables in this approximation, and that there is no need for an equilibration phase to adjust the acceptance rate of each variable. In practice, anharmonic character increases as the eigenvalues decrease even for thermally acceptable moves,[16] and Noguti and Go have suggested scaling the collective variables with eigenvalues below $\lambda_0 = 0.003$ kcal/mol per degree by $\lambda_0^{1/2}$ instead of $\lambda_i^{1/2}$. We adopt their procedure here. Similarly, we select the normal deviates for changes in the scaled collective variables from a Gaussian probability with zero mean and a standard deviation of $\sigma = 1.2$.[16] To this end, a uniform random deviate, x, in the range $(0 \cdots 1)$ is converted in $\Delta\eta$ by the procedure of Box and Mueller.[29] Taking into account the anharmonicity of the collective variables when fixing the average stepsizes (by calculating the actual energy cost of moving in these directions) did not improve sampling).

At the start of the SCV simulation, the test oligomer is minimized, and the second derivative matrix of the potential energy function is calculated.[30] A positive definite Hessian matrix is easily obtained in an internal coordinate system using a quasi-Newton minimization algorithm, which typically yields minimum energy gradients below 10^{-5} kcal/mol per degree. Diagonalization of this matrix provides the required eigenvectors and eigenvalues. For our test oligomer, the eigenvalues range from 10^{-3} to 10^{3} kcal/mol per degree, most lying below 1 kcal/mol per degree. These values cover a range similar to those previously obtained for globular proteins.[16]

The Hessian matrix can be updated during simulation at a chosen step, following a new energy minimization. The simulation can then proceed from the nonminimized structure at this step, thus avoiding a heating phase. However, it should be noted that updating the Hessian upsets the detailed balance conditions imposed on the simulation, so that the subsequent trajectory must be considered as a new MC run. The effect of updating the Hessian is discussed in what follows.

CONFORMATIONAL ANALYSIS

Different conformations of the oligomer sampled during the MC trajectory are analyzed with CURVES[31] to obtain a full helical description including an optimal curvilinear axis. Overall conformations are compared using RMS values based on the atomic coordinates (in Å) or on the angular variables describing the oligomer (in degrees).[32] Thymine methyl rotations are excluded from the latter measurements because they vary considerably, but do not influence overall conformation.

Results

We have carried out 250,000 steps of SV and SCV Monte Carlo simulations at 300 K on the B-DNA form of the 11-mer:

$$\begin{array}{cccccccccccc} G_1 & T_2 & A_3 & T_4 & A_5 & T_6 & A_7 & A_8 & A_9 & A_{10} & C_{11} \\ C_{22} & A_{21} & T_{20} & A_{19} & T_{18} & A_{17} & T_{16} & T_{15} & T_{14} & T_{13} & G_{12} \end{array}$$

followed by several 350,000- or 500,000-step SCV simulations at 400 K using both B and BAB starting conformations. We recall that the BAB form has C_3'-endo sugar puckers associated with all but the 3′-nucleotides of the TATATAAA binding site in the center of the oligomer.[3, 19, 20] Structures were saved every 100 steps for conformational analysis. Computer timings for this oligomer showed that 100,000 steps required roughly 18 hours using a Linux-based Pentium Pro computer. Gains of roughly 30% could be made using a nucleotide-based cutoff for nonbonded interactions, but this was not done for the simulations discussed next.

ACCEPTANCE RATES

For the SV simulation, the maximum stepsizes were adjusted to obtain an acceptance rate close to 0.5. For the SCV simulations, the overall acceptance rate at 300 K varied between 0.35 (B form) and 0.40 (BAB form). Acceptance rates are generally higher for the BAB form, which is probably linked to an increase in overall flexibility. Acceptance rates per variable lie between 0.1 and 0.7 and, for the B form, there is a pronounced trend to higher acceptance rates with increasing eigenvalues. The same trend appeared in calculations performed on a different sequence in the B form (results not presented), but interestingly, not the BAB form. In this case, collective variables with low eigenvalues have acceptance rates almost twice

as high as those for the B form. It is recalled that these steps should allow the largest moves in the harmonic approximation, but that they also present the highest degree of anharmonicity. Calculations have shown that 13 collective variables out of 181 for the BAB form and 28 variables for the B form account for half of the total angular fluctuations. In both cases, 13 variables are associated with eigenvalues below $\lambda_0 = 0.003$ kcal/mol per degree and 76 (BAB form) and 86 (B form) variables account for 90% of the total movement (see Fig. 1). It is also interesting to note that the acceptance rate per variable, averaged every 50,000 steps, shows no monotonic change (Fig. 2). This suggests that the initial Hessian matrix continues to represent the collective movements of the evolving oligomer conformation and that there is no need to regularly update this matrix for simulations of this length.

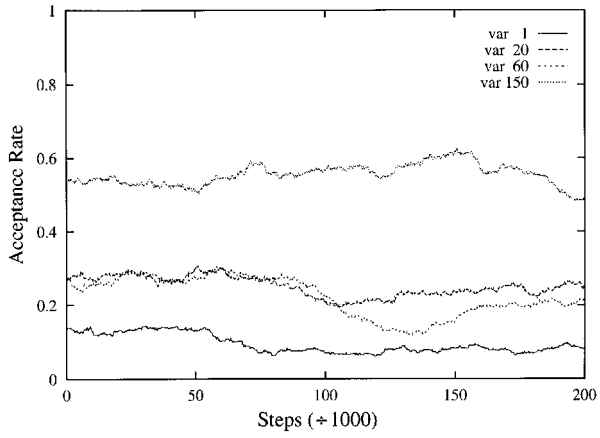

FIGURE 2. Evolution of the acceptance rate (averaged using a 50,000-step window) during the SCV simulation for chosen collective variables. Note that the Hessian matrix is not updated during the simulation.

CONVERGENCE AND SAMPLING EFFICIENCY

The SV and SCV methods exhibit strikingly different behavior in terms of convergence and overall conformational change during the simulation. The evolution of the total energy (Fig. 3) shows that the SCV calculation converges rapidly to a mean value of roughly −430 kcal/mol in less than 20,000 steps. In contrast, the SV energy increases very slowly and does not converge before roughly 200,000 steps.

The degree of conformational change occurring during the simulations can be judged from Figure 4, which shows the angular RMS difference between the initial conformation and the snapshots taken from the simulation. Comparisons are made for complete oligomer and for its central four-nucleotide-pair TATA tetramer. These results confirm the rapid convergence of the SCV method. They also show considerably larger movements in the case of the SCV simulation, notably for the central tetranucleotide. This fragment reaches an RMS of 16° from the initial conformation, as opposed to only 8° in the case of the SV method. The SCV approach thus leads to virtually as much movement in the middle of the molecule as for the whole oligomer and, as hoped, overcomes the steric

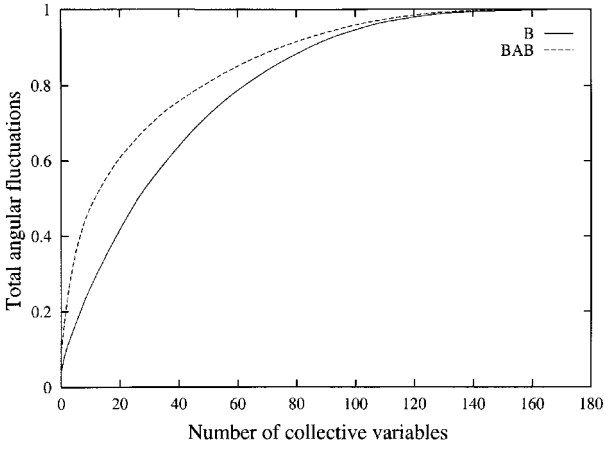

FIGURE 1. Total angular fluctuation of the DNA oligomer (scaled to a maximal value of 1) as a function of the number of collective variables taken into account. Results are shown for the B and BAB starting conformations of the oligomer.

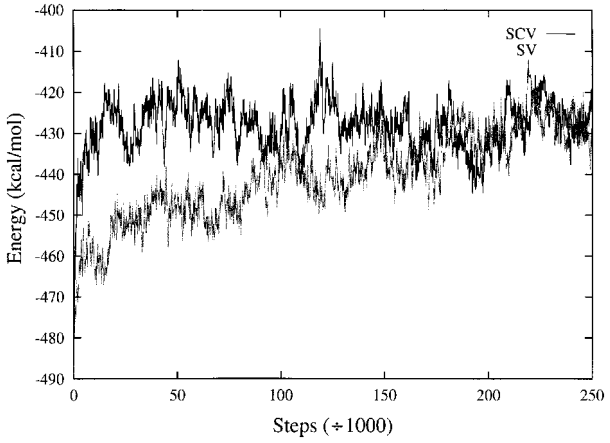

FIGURE 3. Energy variation (kilocalories per mole) during 250,000 steps of the SV and SCV simulations of the test oligomer.

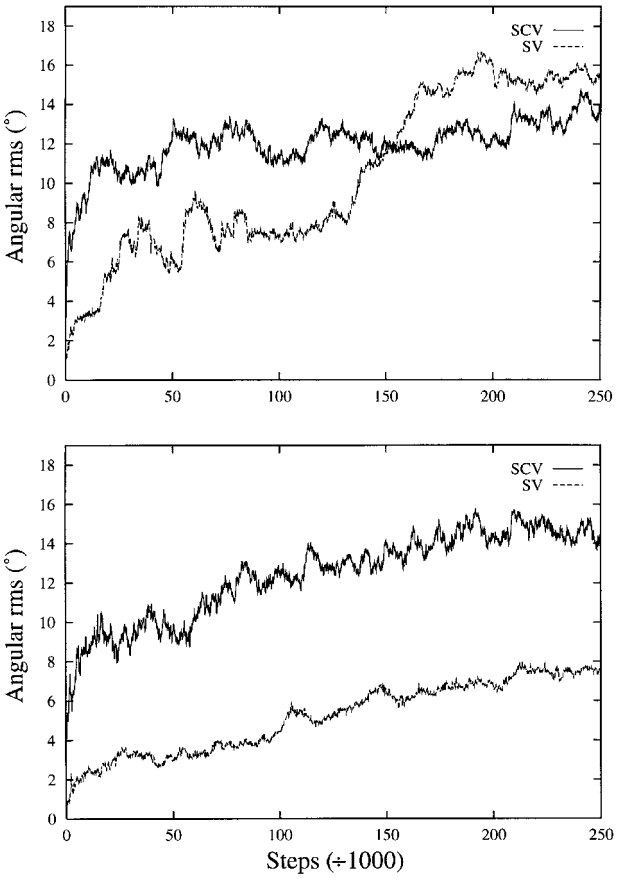

FIGURE 4. Angular RMS differences (degrees) between the initial conformation of the test oligomer and the conformations attained during 250,000 steps of the SV and SCV simulations. The upper portion refers to the whole oligomer, and the lower portion to the central TATA tetramer.

clashes that hinder the SV approach. Given this improvement in sampling, we limit further discussion to results from the SCV-MC approach.

Several other questions can be raised concerning the use of collective variables. First, because these variables are derived from the Hessian matrix describing a given energy minimum, they will restrict the exploration of the conformational space to a region close to this minimum. Monitoring variations during simulation shows that, while variables with higher eigenvalues oscillate quasi-periodically around an equilibrium value, in accord with their harmonic character, variables with low eigenvalues ''escape'' from such behaviour after roughly 20,000 steps and lose track of their initial values. This escape was categorized by calculating a new Hessian matrix after minimizing a structure near the end of the simulation. It was found that less than a quarter of the eigenvectors of the new matrix showed more than 50% identity with corresponding vectors from the original matrix.

We investigated the effects of updating the Hessian matrix in two 500,000-step simulations at 400 K on the B form of our test oligomer. In the first simulation, updates were made three times at 150,000-step intervals. In this case, the last two updates had no impact because the minimized structure returned to the original energy well. In a second simulation, updates were made automatically each time one quarter of the collective variables with eigenvalues below $\lambda_0 = 0.003$ kcal/mol per degree underwent a displacement more than 50° from their initial positions (at 89,700, 116,800, 188,900, 224,100, and 303,300 steps). In both cases, updating the Hessian changed the course of the simulation, but did not result in significant differences in the RMSD reached with respect to the initial conformation. We conclude, in line with the results of the generally stable acceptance rates of the collective variables, that updating is not necessary in runs of the length studied here.

CONFORMATIONAL CHANGES

The B and BAB forms of the 11-mer are now compared in terms of overall and helical fluctuations to examine the role of the latter in forming a complex with TBP. Simulations were carried out with two values of D_0 (1 and 4), the starting value of the sigmoidal dielectric function. The latter value was tested because it leads to more realistic base-pairing energies in aqueous solution and could be expected to have an influence on the overall flexibility of DNA.[26,33] This change had only a minor impact on the fluctuations of individual helical parameters, but did increase overall flexibility, as judged from fluctuations in groove width and axis bending. We have thus retained these simulations (consisting of 350,000 steps for the B and BAB forms) for the discussion that follows.

Table I lists the average values and standard deviations of some important helical parameters over the last 300,000 steps of the simulations on both the B and BAB forms of our oligomer. Data are reported for the two central steps, A5pT6 and T6pA7, as well as averaged values for the central 7 base pairs.

These data give deeper insight into the nature of the conformational fluctuations taking place. First, base-pair opening values are a useful guide to the stability of the base pairing. These parame-

TABLE I.
Mean Values and Standard Deviations (in Parentheses) of Chosen Helical Parameters and of Groove Widths in the Center of the Oligomer.

	A5pT6		T6PA7		Average Over Central 7 bp	
B						
Twist	31.9	(3.6)	33.9	(5.9)	34.4	(1.3)
Rise	3.6	(0.4)	3.3	(0.6)	3.5	(0.2)
Roll	−4.7	(5.5)	2.0	(7.3)	−2.9	(4.4)
Tilt	1.2	(4.7)	−2.5	(7.7)	−1.1	(1.5)
Opening	−1.9	(8.5)	15.9	(15.4)	6.6	(7.9)
Minor groove	6.0	(1.5)	5.2	(1.6)		
Major groove	13.5	(4.2)	12.8	(4.1)		
BAB						
Rise	2.9	(0.6)	3.6	(0.7)	3.2	(0.3)
Twist	26.7	(4.3)	27.0	(5.7)	27.9	(1.8)
Roll	3.9	(8.4)	22.0	(7.2)	10.7	(2.9)
Tilt	1.1	(7.0)	3.9	(7.2)	0.4	(1.4)
Opening	3.1	(13.1)	5.0	(14.4)	6.1	(5.3)
Minor groove	11.7	(0.8)	11.5	(0.7)		
Major groove	10.9	(2.2)	10.9	(2.6)		

ters show that the double helix remains stable, but flexible. For the B form, values oscillate between −20° and 30° with a transitory peak at 60° for T6pA7. Variations in twist are considerable. The increased flexibility of the TpA step, T6pA7, compared with the ApT step, A5pT6, can be judged both from the standard deviations shown in Table I and from the time series shown in Figure 5 for the B-form simulation. Figure 5 also illustrates that helical parameter fluctuations are generally composed of low amplitude rapid movements superimposed on larger amplitude and slower changes. Roll and tilt angles show fluctuations of roughly similar amplitudes, but roll values generally show larger mean values, as expected.[34–36] A particularly high mean roll (22°) is seen for the T6pA7 step within the BAB form of the oligomer, in line with our earlier modeling[20] and with deformations seen within the crystallographic complex of TBP.[19]

Very similar fluctuations to those shown in Table I have been obtained recently in an MD simulation on an oligomer containing the same sequence and using an explicit water and counterion environment.[3] Thus, for the B form of the oligomer, the standard deviation for the twists of A5pT6 and T6pA7 were 4.0° and 5.3°, respectively, whereas rises fluctuated by roughly 0.35 Å, rolls by 4.6° and 7.5°, and tilts by 4.3° and 4.6°, respectively. This overall agreement with the SCV-MC simulation, despite our use of a simplistic solvent model, is encouraging.

In terms of sugar puckering, the B-form simulation shows many transitions from the original C2′-endo puckers to O4′-endo, and, more rarely, to C3′-endo forms. The latter puckers generally occur at the ends of the oligomer, but are also observed for T4 and T16. Such transitions typically require roughly 100,000 steps but fast transitions were observed for the BAB form, both for terminal nucleotides and at the junctions between the B and A-like domains of the oligomer. However, all sugars initially having C3′-endo puckers in the BAB simulation retained this conformation (with the exception of occasional passages to O4′-endo and a single passage to C2′-endo for T2). A plot of phase versus amplitude for A7 and T16 during the B-form simulation (Fig. 6) clearly shows that these pseudorotational variables vary in an uncorrelated manner, emphasizing that they need to be treated as independent variables within an internal coordinate nucleic acid model.

In more macroscopic terms, it was found that our oligomer has a tendency to bend during the simulation, notably in the case of the BAB form. Such fluctuations (measured between the terminal helical axis vectors calculated with CURVES) reach values up to 60° (Fig. 7). Moreover, there is a strong anisotropy in the direction of bending, fa-

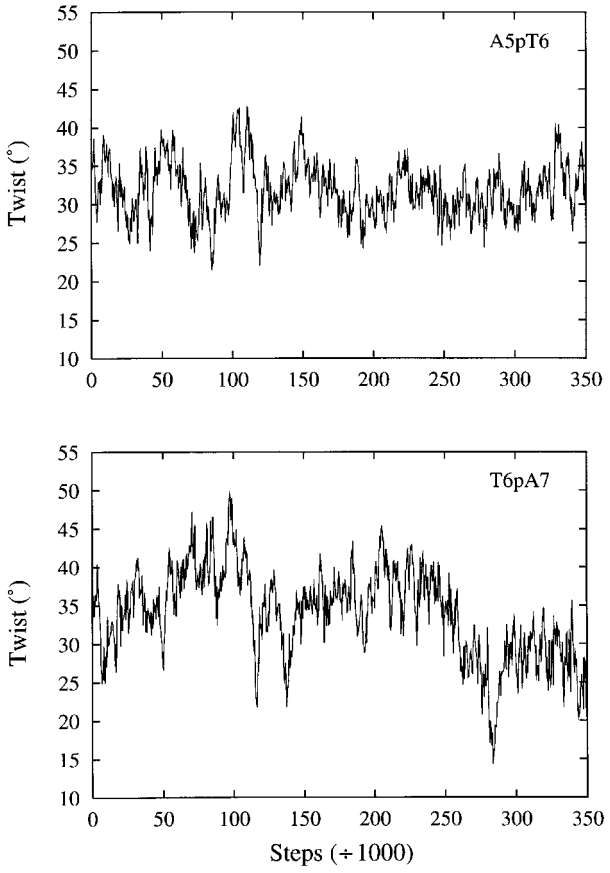

FIGURE 5. Helical twist fluctuations (degrees) during the SCV simulation of the B forms of the oligomer for the A5pT6 and T6pA7 steps.

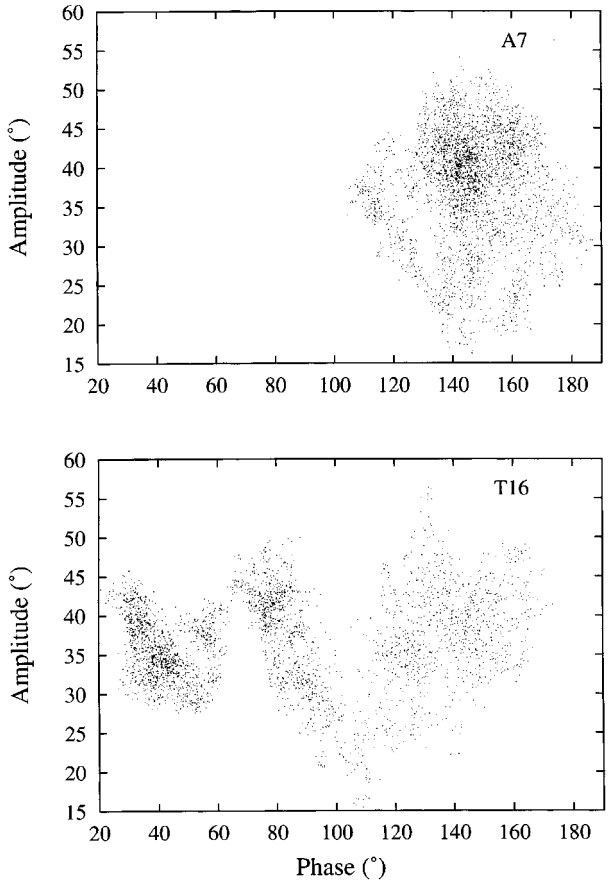

FIGURE 6. Variations in sugar pucker using phase (degrees) versus amplitude (degrees) for the central nucleotides during the SCV simulation of the B form of the oligomer.

voring bending toward the major groove in the center of the oligomer. This result is also similar to that seen by MD simulation.[3] This behavior appears to be linked to the stability of TBP binding[37] and it is again encouraging that it is reproduced by our SCV-MC simulation.

SUBSTATE ANALYSIS

An alternative way to judge how the oligomer is moving through conformational space is to investigate whether it is exploring distinct energy wells. To this end, we have thus energy minimized more than 400 conformations regularly spaced along the simulations (every 800 steps) and clustered the resulting structures on the basis of their angular RMS differences. To distinguish end effects from more global changes, clustering was carried out both on the conformation of the whole 11-mer and on the conformation of the central 7-mer.

The results in Figure 8 illustrate the occupation of the various energy substates for the simulation on the B form, numbered along the vertical axis in order of their appearance (following energy minimization at a given moment in the simulation). We can see from this figure that the simulation remains close to its initial substate for a period of roughly 20,000 steps, in agreement with the RMS fluctuations discussed previously (see Fig. 4). The conformation then begins to change more radically and passes into the regions around a steadily increasing number of substates. Some of these substates are revisited frequently, whereas others occur only fleetingly.

The substates differ from one another by angular RMS values between 5° and 24° and in energy

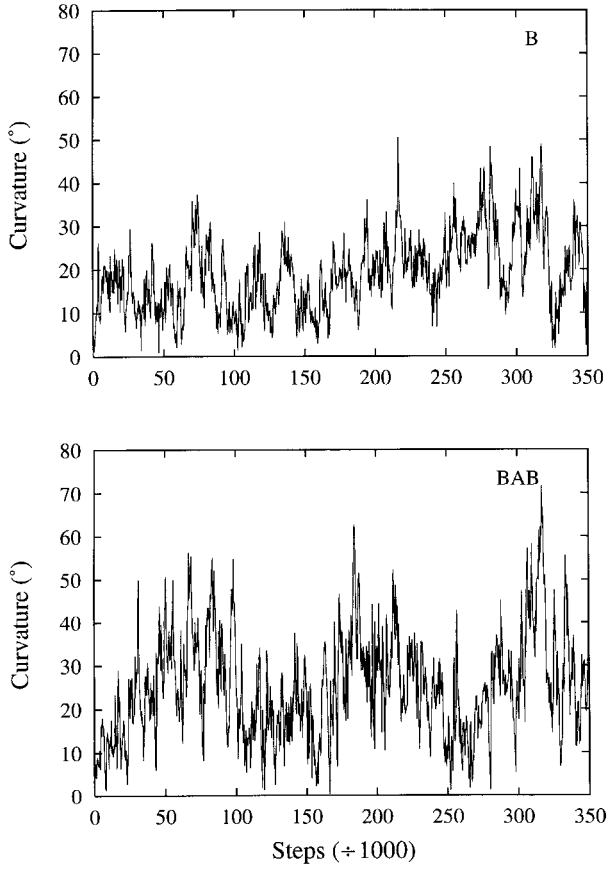

FIGURE 7. Fluctuations in curvature during the SCV simulation on the B and BAB form of the oligomer.

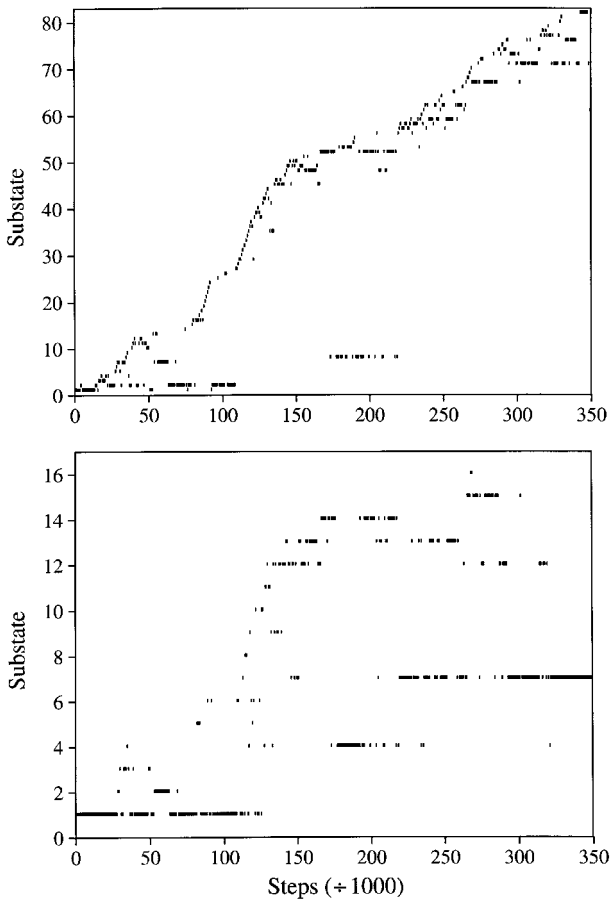

FIGURE 8. Conformational substates visited during the SCV simulation. The vertical axis numbers the substates in order of appearance and the horizontal axis indicates the number of simulation steps carried out. The upper diagram shows the substates for the complete oligomer, whereas the lower diagram refers to the central 7-mer.

by up to 10 kcal/mol (although three fourths of the substates fall within a 5-kcal/mol interval). By the end of the simulation, 82 distinct substates have been visited for the full oligomer in the B form, whereas 43 are found for the central 7-mer. This difference reflects the greater mobility of the terminal nucleotides. It is remarked that each substate can be uniquely characterized by its sugar puckers, as noted in our earlier adiabatic studies of helically symmetric DNA polymers.[11,27] This finding explains why only six substates were located for the central 7-mer of the BAB form, whose sugars are much less flexible (results not shown). Interestingly, this lack of repuckering does not impede large concerted moves in the structure producing the considerable bending fluctuations described.

Conclusions

We have carried out Monte Carlo simulations for nucleic acid oligomers using an internal coordinate model, including a pseudorotational sugar representation and collective variable moves. It is demonstrated that the scaled collective variable approach developed by Noguti and Go is considerably more efficient than single variable moves in the case of a double-stranded DNA oligomer. The results obtained from our simulations show large conformational fluctuations with almost as much movement in the center of the oligomer as at its ends. Sugar repuckering is infrequent, but important fluctuations are seen in all helical parameters and the oligomer undergoes major anisotropic bending, notably in the case of the BAB conformation, in line with results obtained recently using MD simulations in water. The analysis of our simulations also shows that many distinct energy substates are visited during 500,000-step simulations.

Although improvements can clearly be made to the SCV-MC method, both in terms of computational efficiency and in terms of the quality of the environmental model, the results obtained here suggest that this development is worth continuing and can lead to an interesting alternative to MD simulations for studying large conformational fluctuations in nucleic acids.

References

1. T. E. Cheatham III and P. A. Kollman, *J. Mol. Biol.*, **259**, 434 (1996).
2. M. A. Young, B. Jayaram, and D. L. Beveridge, *J. Am. Chem. Soc.*, **119**, 59 (1997).
3. D. Flatters, M. A. Young, D. L. Beveridge, and R. Lavery, *J. Biomol. Struct. Dynam.*, **14**, 757 (1997).
4. V. B. Zhurkin, Y. P. Lysov, and V. L. Ivanov, *Biopolymers*, **17**, 377 (1978).
5. K. J. Miller, *Biopolymers*, **18**, 959 (1979).
6. H. Sklenar, *Studia Biophys.*, **93**, 175 (1983).
7. A. Amadei, A. B. Linssen, and H. J. C. Berendsen, *Proteins*, **17**, 412 (1993).
8. S. H. Northrup and J. A. McCammon, *Biopolymers*, **19**, 1001 (1980).
9. M. Ghomi, J.-M. Victor, and C. Henriet, *J. Comput. Chem.*, **15**, 433 (1994).
10. V. B. Zhurkin, N. B. Ulyanov, A. A. Gorin, and R. L. Jernigan, *Proc. Natl. Acad. Sci. USA*, **88**, 7046 (1991).
11. R. Lavery, in *DNA Bending and Curvature*, W. K. Olson, R. H. Sarma, M. H. Sarma, and M. Sundaralingam, Eds., Adenine Press, New York, 1988, p. 191.
12. R. Lavery, K. Zakrzewska, and H. Sklenar, *Comp. Phys. Commun.*, **91**, 135 (1995).
13. M. Poncin, B. Hartmann, and R. Lavery, *J. Mol. Biol.*, **226**, 775 (1992).
14. R. Lavery and B. Hartmann, *Biophys. Chem.*, **50**, 33 (1994).
15. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.*, **21**, 1087 (1953).
16. T. Noguti and N. Go, *Biopolymers*, **24**, 527 (1985).
17. H. A. Gabb, C. Prévost, and R. Lavery, *J. Comput. Chem.*, **16**, 667 (1995).
18. M. Tomimoto and N. Go, *J. Phys. Chem.*, **99**, 563 (1995).
19. Y. Kim, J. H. Gieger, J. H. Hahn, and P. B. Sigler, *Nature*, **365**, 512 (1993).
20. A. Lebrun, Z. Shakked, and R. Lavery, *Proc. Natl. Acad. Sci. USA*, **94**, 2993 (1997).
21. K. J. McConnell, R. Nirmala, M. A. Young, G. Ravishanker, and D. L. Beveridge, *J. Am. Chem. Soc.*, **116**, 4461 (1994).
22. R. Elber and M. Karplus, *Science*, **235**, 318 (1987).
23. H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, *Science*, **254**, 1598 (1991).
24. H. Sklenar, J. Bernet, R. Lavery, A. Müller, and W. Schultz *J. Am. Chem. Soc.*, submitted.
25. B. E. Hingerty, R. H. Ritchie, T. L. Ferrell, and J. E. Turner, *Biopolymers.*, **24**, 427 (1985).
26. B. Jayaram, S. Swaminathan, D. L. Beveridge, K. Sharp, and B. Honig, *Macromolecules*, **23**, 3156 (1990).
27. V. Fritsch and E. Westhof, *J. Am. Chem. Soc.*, **113**, 8271 (1991).
28. M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods, Vol. I: Basics*, Wiley, New York, 1986.
29. W. H. Press, S. A. Teukolsky, W. T. Vettering, and B. P. Flannery, *Numerical Recipes in Fortran*, 2*nd Ed.*, Cambridge University Press, Cambridge, UK (1992).
30. T. Ha Duong and K. Zakrzewska, *J. Comput. Chem.*, **18**, 796 (1997).
31. R. Lavery and H. Sklenar, *J. Biomol. Struct. Dynam.*, **6**, 665 (1989).
32. R. Lavery, In *Modelling Biomolecular Structures and Mechanisms* A. Pullman, B. Pullman, and J. Jortner, Eds., Kluwer, Dordrecht, 1995, p. 217.
33. N. Arora and B. Jayaram, *J. Comput. Chem.*, **18**, 1245 (1997).
34. V. B. Zhurkin, Y. P. Lysov, and V. I. Ivanov, *Nucl. Acids Res.*, **6**, 1081 (1979).
35. W. K. Olson, A. R. Srinivasan, R. C. Maroun, R. Torres, and W. Clark, In *Unusual DNA Structures*, R. D. Wells and S. C. Harvey, Eds., Springer, New York, 1989, p. 207.
36. S. R. Sanghani, K. Zakrzewska, S. C. Harvey, and R. Lavery, *Nucl. Acids Res.*, **24**, 1632 (1996).
37. J. D. Parvin, R. J. McCormick, P. A. Sharp, and D. E. Fisher, *Nature*, **373**, 724 (1995).